Short-Term Forecasting of Building Energy Consumption with Deep Generative Learning

Yichuan X. Ma[1]
[1]The University of Hong Kong, Hong Kong SAR, China

Abstract

Short-term forecasting of building energy consumption is highly valuable from both technical and economic point of views. In this paper, a deep generative learning method taking account of short-term future meteorological data is proposed to forecast building energy consumption in the next 24 hours. A conventional multilayer perceptron and a non-meteorology version of the proposed GAN-based model were developed and comparatively tested as baseline models. Multi-year hourly meteorological data and actual energy consumption measurements from two office buildings in Shanghai were used for modelling and testing. The proposed model significantly outperformed the baseline models in all granularity settings. Decent cross-case generalisability of the proposed GAN-based models were demonstrated.

Key Innovations

- A novel generative short-term forecasting framework is delineated.
- Customised conditional GAN-based models with a 1D-UNet generator under the generative short-term forecasting framework are developed and quantitatively evaluated under multiple temporal granularity settings.
- Historical and future meteorological information is taken into consideration to improve forecasting accuracy and the impact of its uncertainty on building energy forecasting is analysed.
- Cross-case generalisability, which is an underappreciated issue in the field of building energy forecasting, is discussed.

Practical Implications

With (or without) access to future meteorological data, practitioners can implement the proposed GAN-Plus (or GAN-Zero) model under the generative short-term forecasting framework to achieve accurate one-day ahead forecasting of building energy consumption with decent cross-case generalisability and high robustness to weather uncertainties.

Introduction

Short-term forecasting of building energy consumption, though challenging, is highly valuable from both technical and economic point of views (Deb et al. 2017). The information-intensive nature of modern building energy management system has made building energy forecasting possible via inverse modelling methods such as deep learning. The generative adversarial network (GAN) (Goodfellow et al. 2014) is one of the most successful techniques in the field of deep learning in recent years. Its main concept is to facilitate learning through the zero-sum contest between two neural networks, where one network is trying to discern the samples generated by the other network from the real ones, and the other network is trying to generate real-like samples from noises to fool its opponent.

Though some efforts were taken to make indirect application of GAN for building energy forecasting, such as using the discriminator of GAN as a feature extractor (Fan et al. 2019) or employing GAN as a tool for data augmentation (Tian et al. 2019), a direct application of such generative learning techniques still remains a challenge, given the forecasting problem is not naturally well-fit with the original context of GAN. Besides this major gap, a few more gaps have been noticed. First, The cross-case generalisability (i.e. the ability of the model to extrapolate over new buildings) remains largely underappreciated in the field of building energy forecasting. Second, similar to those in forward modelling (e.g. whole building energy simulation), meteorological data provide a bottom-up pathway for a computational algorithm to build up the model about building energy consumption and are thus expected to play a role in improving the accuracy of the forecasting model, while full exploitation of meteorological data in inverse modelling has yet to be developed. Furthermore, some previous building energy forecasting studies focus on computed energy load/demand instead of measured energy consumption, which may lead to discrepancies with the actual circumstances. The above-mentioned gaps are tackled in this study.

Methodology

Formulation of the forecasting problem

The building energy forecasting problem can be formulated as finding the mapping from the prior energy series and other features to future energy series, i.e. find f: $\{E_{prior}, S\} \rightarrow E_{forecast}$. The forecasting target is the future building energy consumption $E_{forecast}$. E_{prior} refers to the historical building energy consumption, and S refers to other features, which may include the historical and future meteorological data, holiday information, etc.

A generative framework for short-term forecasting

Here we present a novel patch-generating framework for short-term forecasting. Rather than passing the data at single time points to the model, we apply a sliding time window to the data to generate multiple segments (See detailed parameter settings in the "experimental settings of the case study" section). Each segment of the energy time series comprises two parts, namely a piece of prior information with a width of w_{prior} , and a patch of all zeros with a width of $w_{forecast}$. The task of the short-term forecasting model is to replace the zeros with predictions as close to the ground truth as possible. Under this framework, the short-term forecasting is converted into a patch-generation problem, as shown in Figure 1. This framework is compatible with conditional GAN (Mirza and Osindero 2014), which generates samples from given conditions instead of noises.

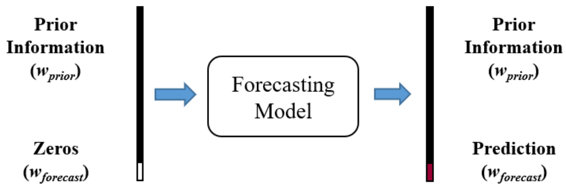

Figure 1: The generative short-term forecasting framework.

Proposed and baseline models

Three forecasting models, namely MLP (baseline 1), GAN-Zero (baseline 2) and GAN-Plus (proposed), were designed, implemented and investigated in this study. Besides, a random guess model was also implemented to provide the chance level performance as a general baseline.

MLP: First, a conventional machine learning model, namely a multilayer perceptron (MLP, also known as a dense neural network), was implemented as one of the baseline models. Compared with traditional time-series based techniques such as the autoregressive integrated moving average model (ARIMA), MLP is reported to have a better performance in the task of forecasting due to its capability in capturing nonlinear patterns (Aslanargun et al. 2007). The architecture of MLP is shown in Figure 2. Each hidden layer (blue block, 384 neurons/1 channel for each layer) is fully connected with its adjacent layers and is followed by a rectified linear unit (ReLU) activation function to introduce nonlinearity.

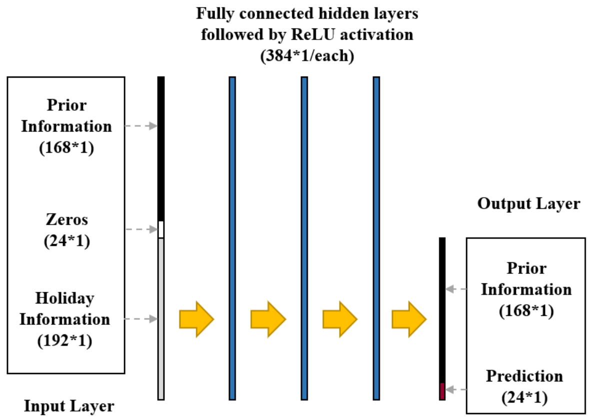

Figure 2: A schematic of the architecture of MLP.

GAN-Zero and GAN-Plus: Two models were implemented based on GAN, including a GAN without meteorological inputs (GAN-Zero), and a GAN with meteorological inputs (GAN-Plus). Both of these two models use a 1D-UNet as the generator and a deep convolutional neural network (CNN) as a classifier. The only difference between GAN-Zero and GAN-Plus is the number of hidden channels (i.e. the number of convolutional kernels).

Figure 3 shows the general architecture of the 1D-UNet generator. It was modified from the original 2D-UNet (Ronneberger et al. 2015), which was first introduced for medical image segmentation and has a large number of successful applications in the field of computer vision. The generator is composed of 4 contracting blocks (the left half) and 4 expanding blocks (the right half), with an upsampling layer at the beginning and a downsampling layer at the end. Short cut pathways (orange arrows) were implemented between corresponding layers. For simplicity, the components were drawn in blocks and not to scale here. The loss of the generator was composed of two components, namely an adversarial loss from the discriminator and a reconstruction loss from comparing the predicted outputs with the ground truth. The contribution of each component was modulated correspondingly by a hyperparameter λ, as shown in (1).

$$Loss_G = Loss_{adv} + \lambda Loss_{recon} \quad (1)$$

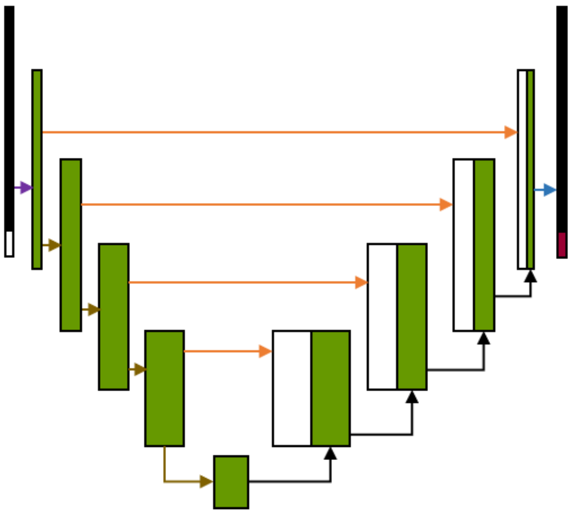

Figure 3: A schematic of the architecture of the 1D-UNet generator of GAN-Zero and GAN-Plus.

The discriminator was a 4-layer CNN which took the generated samples and true samples as inputs and outputted a binary value indicating that the input was real or fake (i.e. generated by the generator). The loss of the discriminator was defined as its classification accuracy. The adversarial loss of the generator, on the other hand, was defined as the rate of how often the discriminator made mistakes.

Figure 4 shows the design of the inputs for GAN-Zero and GAN-Plus. Holiday information (0 for workdays, 1 for Sundays and public holidays, and 0.75 for Saturdays) was combined with the historical building energy consumption data (followed by a patch of zeros) for both models. Note that the holiday information is not rigorously periodic, as the public holidays are sparsely and irregularly distributed over time. For GAN-Plus, the meteorological data (orange blocks) were channel-wisely concatenated with the inputs. Due to the nonlinearity of the coupling between building energy consumption and meteorological parameters (Ma and Yu 2020), it is important to include the range of the meteorological parameters (mainly the temperature) in the inputs to ensure efficient information utilisation (the yellow block). Generally, the input sizes of GAN-Zero and GAN-Plus are 192 units/2 channels and 192 units/7 channels respectively.

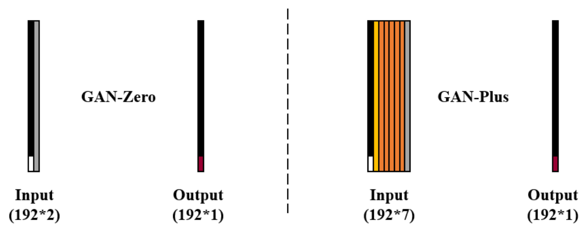

Figure 4: Design of the inputs for GAN-based models.

Chance Level: For the random guess model, 10000-fold Monte Carlo simulations were conducted to calculate the chance level performance.

Experimental settings of the case study

Data used in this study include (1) actual building energy consumption data measured from a medium-rise office building (B1) in Shanghai during 2015-2017, (2) actual building energy consumption data measured from a high-rise office building (B2) in Shanghai during 2017, (3) actual meteorological data in Shanghai during 2015-2017, including dry-bulb air temperature, dew temperature, relative humidity, air pressure and wind speed, and (4) holiday information.

Figure 5 shows the processes of model training and testing. The data from B1 during 2015-2016 (training set, including more than 17000 observations) were used for training, and the data from B1 during 2017 (testing set 1, including more than 8000 observations) as well as the data from B2 during 2017 (testing set 2, including more than 8000 observations) were held out for testing. After the training process, the trained models were tested on the two distinct testing sets for different purposes. On testing set 1, the model was evaluated in terms of forecasting performance. On testing set 2, the model was evaluated in terms of cross-case generalisability by comparing the performance with it on testing set 1. Note that there is no overlap between the training set and the testing sets.

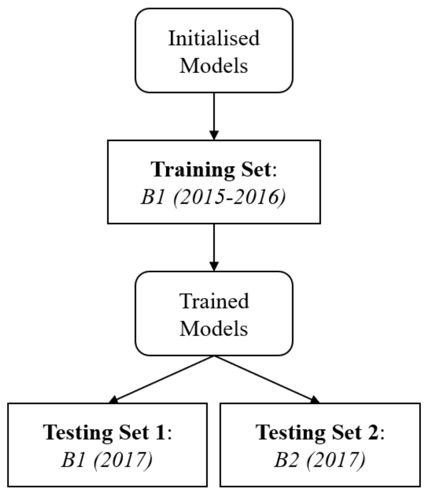

Figure 5: Model training and testing.

The coefficient of variation of the root-mean-square error (CV-RMSE) of the forecasted building energy consumption was used as the evaluation metric of forecasting performance, defined as (2). Superior performance (higher accuracy) of the model is indicated by a lower CV-RMSE. Compared with other metrics, CV-RMSE has advantages in measuring the normalised accumulated error and therefore is expected to better reflect the general accuracy of the model (Royapoor and Roskilly 2015). When n equals 1, the CV-RMSE is equivalent to the root mean square error (RMSE) and mean absolute error (MAE).

$$CV_{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - \widehat{y_k})^2} \Big/ (\frac{1}{n}\sum_{k=1}^{n} y_k) \qquad (2)$$

The spectral density estimation method (Lomb 1976) was used to derive the intrinsic periodogram of the building energy consumption data. As shown in Figure 6, the most prominent frequencies are approximately 1/24 h^{-1} and 1/168 h^{-1}, which correspond to the daily and weekly periodicities, respectively. A width of 168 hours was thus selected to give prior information in a full period for the model. The width of the predicted time series and the step size of the sliding window for segment generation were both selected as 24 hours.

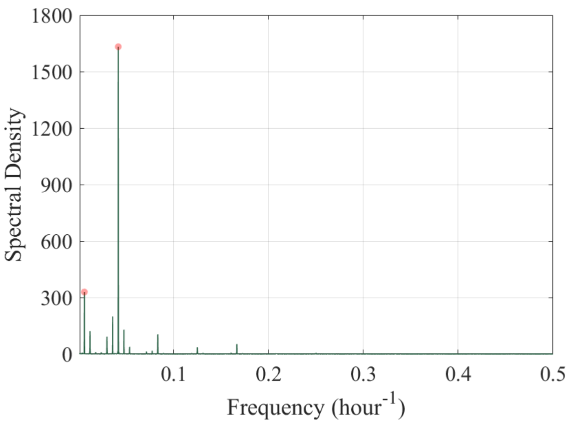

Figure 6: Periodogram power spectral density estimation of the building energy consumption data.

In order to test the robustness of the GAN-Plus model, noises were injected into the future meteorological inputs by masking random positions of the inputs with zeros. The robustness test was conducted on the testing set 1.

Results and Discussion

Statistical profiles of energy and meteorological inputs

The relationships between building energy consumption of a case building (B1) and meteorological inputs during the three continuous natural years are shown in Figure 7. Nonlinear relationships between building energy consumption and temperature and distinct clusters of weekdays and holidays/Sundays/Saturdays were observed, which is consistent with previous findings with multi-year simulated building energy data (Ma and Yu 2020). In general, building energy consumption is negatively (positively) correlated with dry-bulb temperature during the winter half-years (summer half-years). Temperature changes in different value ranges can cause discrepant effects on building energy consumption. Such patterns can be more clearly found in coarser granularity settings (Figure 7C). These results confirm the importance of considering meteorological features and their uncertainties in building energy forecasting.

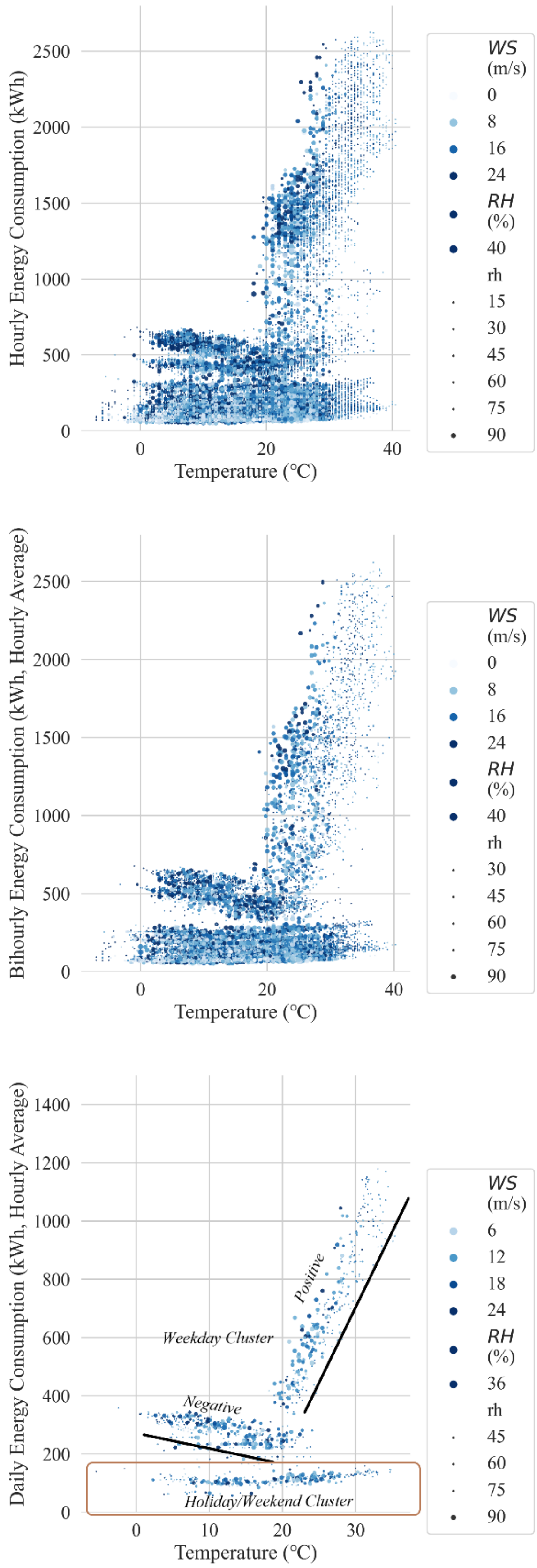

Figure 7: Statistical profiles of energy and meteorological data in (A, top) hourly, (B, middle) bihourly and (C, bottom) daily granularity settings. RH: relative humidity. WS: wind speed.

Forecasting performance

The building energy consumption forecasting performance of the proposed and baseline models was evaluated on the testing set 1, about which the models had no prior information. GAN-Plus constantly achieved higher accuracies compared with the baseline models (Table 1).

Table 1. Performance (CV-RMSE of forecasted energy time-series) of the proposed and baseline models on the testing set 1.

Model	Granularity		
	Hourly	Bihourly	Daily
MLP	20.40%	19.53%	12.57%
GAN-Zero	15.61%	13.98%	6.13%
GAN-Plus	**14.52%**	**12.26%**	**5.23%**
Chance Level	78.86%	77.99%	60.60%

Conceivably and intuitively, a coarser granularity could lead to higher accuracy, and it is confirmed by the results.

Cross-case generalisability

As shown in Table 2, MLP failed in the cross-case forecasting task with an hourly CV-RMSE notably larger than 30%. Though similar in terms of hourly forecasting performance, GAN-Plus showed superior accuracy (with a difference > 1%) than GAN-Zero in daily forecasting. This could be due to the stronger correlation between building energy consumption and meteorological data with reduced noise under the coarser granularity settings. Both GAN-Plus and GAN-Zero displayed the capability of extrapolating over temporal & cross-case conditions that have not been previously seen/measured.

Table 2. Performance (CV-RMSE of forecasted energy time-series) of the proposed and baseline models on the testing set 2.

Model	Granularity		
	Hourly	Bihourly	Daily
MLP	46.57%	43.06%	27.01%
GAN-Zero	22.35%	21.35%	8.13%
GAN-Plus	**22.23%**	**20.81%**	**7.05%**
Chance Level	78.10%	77.94%	58.84%

Robustness to noises in future meteorological data

In the robustness test, GAN-Plus demonstrated strong capability in dealing with the uncertainties in the future meteorological clues, with less than 1.5% deterioration of accuracy with a noise level of 20% (Table 3). Note that when the noise level of future meteorological data reached 20%, the accuracy was lower than the non-meteorological model. This indicates the potential risk of using the GAN-Plus model when the future meteorological data is not reliable. Empirically, the accuracy of weather forecasting decreases with the target predicted period, which is approximately 90% (85%) for 1 to 3 (4 to 7) days ahead. Thus, an improvement over the GAN-Zero model is normally expected/guaranteed. When accurate weather forecasting is not guaranteed, it is recommended to use the GAN-Zero model instead.

Table 3. Noise robustness of GAN-Plus.

Noise Level	**0%**	**10%**	**20%**
CV-RMSE	14.52%	15.43%	15.96%

Overall Discussion

The primary purpose of this work is to explore whether GANs can enable one-day ahead forecasting of building energy consumption with sufficient accuracy for real-world applications. In general, according to the ASHRAE standard (Monfet and Radu Zmeureanu PhD 2009), a building energy model is calibrated if its hourly CV-RMSE is less than 30% / monthly CV-RMSE is less than 15%. In this case, all the MLP, GAN-Zero and GAN-Plus fulfilled the requirement in the individual forecasting task. However, when forecasting in a cross-case setting, MLP failed. Both GAN-based models, on the other hand, still worked.

There are several distinct objectives for the comparisons between the proposed and baseline models. First, the chance level performance provides a general baseline to examine whether other models are truly solving the forecasting problem. Second, the MLP model offers a baseline in terms of the traditional forecasting method, and the comparison between MLP and GAN-Zero evaluates whether GAN has a stronger ability over the plain deep neural network in the time-series forecasting task under the newly proposed generative framework. Third, the proposed GAN-Plus aims to explore the potential room for improvement by exploiting the meteorological information based on GAN-Zero.

The proposed customised GAN-based generators displayed superiority in serving as forecasting models due to their strong capability in grasping the in-depth features of the data through their adversaries (i.e. the discriminators), which was demonstrated through their better performance according to the results. One disadvantage of the GAN model is that it is relatively more computationally expensive to train. In the current case, the training of MLP ran approximately 20 times faster than the one of GANs. Another limitation of GAN-based forecasting is that GAN is originally designed for image processing; thus, there could be deteriorations in performance when dealing with time-series data.

Prior to this study, GAN was only used as a feature extractor or a data augmentation tool in the field of building energy forecasting. This work provided a novel research pathway under the novel generative forecasting framework to directly taking advantage of the success of deep generative learning to forecast building energy consumption.

Furthermore, this work is based on measured building energy consumption, which can better reflect the actual operation situation and provide more accurate information and more meaningful guidance for practitioners. Given the fact that measured energy consumption is noisier than design energy load/demand (e.g. design cooling load) with respect to the historical information and meteorological clues, higher accuracy is expected to achieve when applying the proposed methods in the scenarios of design energy load forecasting.

This contribution should be regarded as a report on preliminary studies. Future works may include (1) developing the image encoding of the energy consumption data to utilise the success of GAN and UNet by enabling the 2D implementation of the forecasting model, (2) investigating the capabilities of the forecasting models under multiple efficiency settings (i.e. considering forecasting with variant width of the prior and target time series), and (3) exploring feasible approaches to utilise heterogeneous data from a group of buildings. Prospective investigations and results in these directions will be presented elsewhere.

Conclusion

A deep generative learning method taking account of short-term future meteorological data is proposed to forecast building energy consumption in the next 24 hours. A non-meteorology version of the proposed GAN-based model and a conventional multilayer perceptron were implemented and comparatively tested as baseline models. State-of-the-art accuracy and decent cross-case generalisability of the proposed GAN-based models were demonstrated. The main findings are:(1) The customised GAN-based models have superior performance than MLP in short-term forecasting of building energy consumption; (2) By exploiting meteorological information, the forecasting accuracy of GAN-based models can be further improved with strong robustness to the future weather uncertainties; and (3) The proposed GAN-Plus model outperformed the chance level and all the baseline models, achieving accuracies of 85.48% with hourly granularity and 94.77% with daily granularity. Decent cross-case generalisability of the proposed GAN-based models were demonstrated.

References

Aslanargun, A., M. Mammadov, B. Yazici and S. Yolacan (2007). "Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting." Journal of Statistical Computation and Simulation **77**(1): 29-53.

Deb, C., F. Zhang, J. Yang, S. E. Lee and K. W. Shah (2017). "A review on time series forecasting techniques for building energy consumption." Renewable and Sustainable Energy Reviews **74**: 902-924.

Fan, C., Y. Sun, Y. Zhao, M. Song and J. Wang (2019). "Deep learning-based feature engineering methods for improved building energy prediction." Applied energy **240**: 35-45.

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio (2014). "Generative adversarial networks." arXiv preprint arXiv:1406.2661.

Lomb, N. R. (1976). "Least-squares frequency analysis of unequally spaced data." Astrophysics and space science **39**(2): 447-462.

Ma, Y. X. and C. Yu (2020). "Impact of meteorological factors on high-rise office building energy consumption in Hong Kong: From a spatiotemporal perspective." Energy and Buildings **228**: 110468.

Mirza, M. and S. Osindero (2014). "Conditional generative adversarial nets." arXiv preprint arXiv:1411.1784.

Monfet, D. and P. Radu Zmeureanu PhD (2009). "Calibration of a building energy model using measured data." ASHRAE Transactions **115**: 348.

Ronneberger, O., P. Fischer and T. Brox (2015). U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, Springer.

Royapoor, M. and T. Roskilly (2015). "Building model calibration using energy and environmental data." Energy and Buildings **94**: 109-120.

Tian, C., C. Li, G. Zhang and Y. Lv (2019). "Data driven parallel prediction of building energy consumption using generative adversarial nets." Energy and Buildings **186**: 230-243.